Eliminating cycles in data base schemas

Y. Hanatani

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:954 90 20

Rapports de Recherche

N° 178

ELIMINATING CYCLES IN DATA BASE SCHEMAS

350 ... 7.2.83
... 320

Yoshito HANATANI

Décembre 1982

ELIMINATING CYCLES IN DATA BASE SCHEMAS

Yoshito HANATANI

Abstract

Call J-schema (resp. M-schema) a schema which can be described by a single join-dependency (resp. by a set of multi-valued dependencies). Call JM-schemas a schema which is both a J-schema and an M-schema. [FMU] and [BFMY] introduced the notion of an acyclic J-schema and showed that a J-schema is acyclic if and only if it is a JM-schema. Acyclic schemas have certain desirable properties and this led to the general belief that cyclic J-schemas were "bad" schemas. In this paper we show that some cyclic J-schemas are in fact useful.

We introduce the notion of a simple schema. A schema S is simple if there exists an acyclic schema S' such that S and S' satisfy the same set of multi-valued dependencies. The main results in the paper are as follows :
a) There are some cyclic J-schemas which are simple. The cycles contained in such schemas are somehow "superficial" and consequently, simple cyclic J-schemas are as useful as their corresponding acyclic schemas. We formalize the procedure for eliminating "superficial" cycles.
b) The only simple M-schemas are the JM-schemas.
c) We study the relationship between J-schemas and M-schemas, and we obtain an elegant characterization of simplicity (consequently, of acyclicity).

Résumé

Appelons J-schéma (resp. M-schéma) un schéma qui peut être décrit par une seule dépendance de jointure (resp. par un ensemble de dépendances multi-valuées). Appelons JM-schéma un schéma qui soit à la fois un J-schéma et un M-schéma. [FMU] et [BFMY] ont introduit la notion de J-schéma acyclique et ils ont montré qu'un J-schéma est acyclique si et seulement si c'est un JM-schéma. Les schémas acycliques ont certaines propriétés intéressantes. On en a généralement conclu que les J-schémas cycliques sont de "mauvais" schémas. Dans ce papier on montre que certains J-schémas cycliques sont en fait utiles.

On introduit la notion de schéma simple. Un schéma S est simple s'il existe un schéma acyclique S' tel que S et S' satisfassent le même ensemble de dépendances multi-valuées. Les résultats principaux de ce papier sont les suivants :
a) Il y a des J-schémas cycliques qui sont simples. Les cycles contenus dans tels schémas sont plus ou moins "superficiels" et par conséquent, les J-schémas cycliques simples sont aussi utiles que leurs schémas acycliques associés. On formalise le processus pour éliminer des cycles "superficiels".
b) Les seuls M-schémas simples sont les JM-schémas.
c) On étudie la liaison entre des J-schémas et des M-schémas, et on obtient une caractérisation élégante de la simplicité (et par conséquent, de l'acyclicité).

INTRODUCTION

Let U be a <u>finite</u> set of attributes that we shall call the <u>universe of attributes</u>. We suppose that to each attribute x of U is associated a domain D_x of its values, that we shall call the <u>domain</u> of x. The only condition on D_x is : card $D_x \geq 2$. The direct (cartesian) product $\otimes_{x \in U} D_x$ of all these domains is called the set of <u>tuples</u> on U, and a <u>finite</u> set of tuples on U is called a <u>relation</u> on U. The set of all the tuples on U will be denoted by TUPLE(U) and the set of all the relations on U will be denoted by Rel(U).

For any subset X of U, the projection on X of a tuple f on U is defined in the usual way and is denoted by f[X]. The projection on X of a relation R on U is then defined as follows :

$$R[X] =_{\mathrm{def}} \{f[X] \mid f \in R\}.$$

A family $\underline{X}$ of subsets of U whose union $\cup\underline{X}$ is equal to U is called a <u>data base schema on</u> U, or simply, a <u>schema</u> on U.

A data base schema on U is therefore a synonym of a covering of U. The set of all the data base schemas is formally defined as follows :

$$C(U) =_{\mathrm{def}} \{\underline{X} \subseteq \mathcal{P}(U) - \{\emptyset\} \mid \cup\underline{X} = U\}.$$

For any relation R on U and for any data base schema $\underline{X}$, the family of the projections R[X] with $X \in \underline{X}$, is called the <u>decomposition</u> of R according to the data base schema $\underline{X}$. For any decomposition of R according to a schema $\underline{X}$, its <u>join</u> is defined as follows :

$$\bowtie \{R[X] \mid X \in \underline{X}\} =_{\mathrm{def}} \{f \in \mathrm{TUPLE}(U) \mid \forall X \in \underline{X}(f([X] \in R[X])\}.$$

As every R[X] is finite and as $\cup\underline{X} = U$, the above join must be a finite set of tuples on U, i.e. it is a relation on U. But in general, this relation contains more of tuples than the original relation R. Naturally it is desirable that the following condition holds :

$$R = \bowtie \{R[X] \mid X \in \underline{X}\}.$$

Such conditions are expressed by the notion of join dependency.

The set of join dependencies (JD's) on U is defined as a set of expressions $\bowtie \underline{X}$ associated to every data base schema $\underline{X}$ on U, i.e.

$$JD(U) =_{def} \{\bowtie \underline{X} \mid \underline{X} \in C(U)\}.$$

For any $R \in Rel(U)$ and for any $\bowtie \underline{X} \in JD(U)$, we define the relation "$R \models \bowtie \underline{X}$" (R satisfies $\bowtie \underline{X}$) as follows :

$$R \models \bowtie \underline{X} <=>_{def} R = \bowtie \{R[X] \mid X \in \underline{X}\}.$$

As a weaker notion, we have the notion of multi-valued dependency. The set of multi-valued dependencies on U is a set of expressions as follows :

$$MVD(U) =_{def} \{X \twoheadrightarrow Y \mid X \subseteq U, Y \subseteq U\}.$$

For any $R \in Rel(U)$ and for any $X \twoheadrightarrow Y \in MVD(U)$, the satisfaction relation "$R \models X \twoheadrightarrow Y$" is defined as follows :

$$R \models X \twoheadrightarrow Y <=>_{def} R = \bowtie \{R[X \cup Y], R[X \cup (U-Y)]\}.$$

Consequently, we have :

$$R \models X \twoheadrightarrow Y <=> R \models \bowtie \{X \cup Y, X \cup (U-Y)\}.$$

That is, every element $X \twoheadrightarrow Y$ of MVD(U) is representable by the element $\bowtie \{X \cup Y, X \cup (U-Y)\}$ of JD(U). But if the cardinality of U is ≥ 3, then an element of JD(U) cannot be represented by an element of MVD(U) in general. For a given element j of JD(U), there are in general many elements m of MVD(U) such that :

$$\forall R \in Rel(U)(R \models j => R \models m).$$

Let us denote by M(j) the set of all such elements m, i.e. :

$$M(j) =_{def} \{m \in MVD(U) \mid \forall R \in Rel(U)(R \mid\!\sim j \Rightarrow R \mid\!\sim m)\}.$$

Then by definition :

$$\forall R \in Rel(U)\ (R \mid\!\sim j \Rightarrow R \mid\!\sim M(j)),$$

where we denote by $R \mid\!\sim M(j)$ the fact that :

$$\forall m \in M(j)\ (R \mid\!\sim m).$$

That is, j semantically implies M(j).

The <u>semantical implication</u> is defined more generally as follows : Let C_1 and C_2 be any two (sets of) conditions (called integrity constraints) such that the satisfaction relations $R \mid\!\sim C_1$ and $R \mid\!\sim C_2$ are defined for any $R \in Rel(U)$. Then the semantical implication "$C_1 \models_U C_2$" of C_2 from C_1, is :

$$C_1 \models_U C_2 \Leftrightarrow_{def} \forall R \in Rel(U)\ (R \mid\!\sim C_1 \Rightarrow R \mid\!\sim C_2).$$

When there is no fear of confusion on the universe set U, we write "$\models$" instead of "$\models_U$".

For any element j of JD(U) :

$$M(j) =_{def} \{m \in MVD(U) \mid j \models m\}$$

and consequently :

$$j \models M(j).$$

Let j, j' be any two elements of JD(U). Then we may compare them by the inclusion relation between M(j) and M(j'). Thus we obtain a partial order on the set JD(U). That is :

$$j \leq_{MVD(U)} j' \Leftrightarrow M(j) \subseteq M(j').$$

Generalizing it, we can define following two relations between any two (sets of) integrity constraints C_1 and C_2 of relations on U :

$$C_1 \leq_{MVD(U)} C_2 \Leftrightarrow_{def} \forall m \in MVD(U)(C_1 \models m \Rightarrow C_2 \models m);$$

$$C_1 \leq_{JD(U)} C_2 \Leftrightarrow_{def} \forall j \in JD(U)(C_1 \models j \Rightarrow C_2 \models j).$$

When there is no fear of confusion on the universe U, we denote them as "$C_1 \leq_{MVD} C_2$" and "$C_1 \leq_{JD} C_2$", respectively. Like as in the case of usual order relations, the conventional use of the notations "$\geq_{*D}$", "$=_{*D}$", "$<_{*D}$" and "$>_{*D}$" are supposed defined, for $*D$ = MVD and for $*D$ = JD, respectively.

We note the followings :

Lemma

$$C_1 \leq_{JD(U)} C_2 \Rightarrow C_1 \leq_{MVD(U)} C_2,$$

for any C_1, C_2 of integrity constraints. □

Corollary

$$C_1 =_{JD(U)} C_2 \Rightarrow C_1 =_{MVD(U)} C_2,$$

for any C_1, C_2 of integrity constraints. □

Lemma

For any $j \in JD(U)$ and for any $M \subseteq MVD(U)$, we have :

(1) $j =_{MVD} M \Rightarrow j \geq_{JD} M$,

and

(2) $j \neq_{MVD} M \Rightarrow j \neq_{JD} M$. □

For every subset M of MVD(U), let us denote by J(M) the set of all the elements j of JD(U) such that $j =_{MVD} M$, i.e. :

$$J(M) =_{def} \{j \in JD(U) \mid j =_{MVD} M\}.$$

We call J(M) the set of the <u>join-realizations</u> of M. J(M) can be empty :

<u>Example</u>

Let U_0 be a set of four distinct attributes, say $U_0 = \{x,y,z,v\}$ and let M_0 be the subset $\{x \twoheadrightarrow y,\ y \twoheadrightarrow x\}$ of $MVD(U_0)$. Then for any element j of $JD(U_0)$, the relation $j =_{MVD} M_0$ does not hold, i.e. $J(M_0) = \emptyset$. In fact, let $j = \bowtie \underline{X}$, then the subsets $\{x,y\}$ and $\{z,v\}$ should not be connected in $\underline{X}$, by the fact that $j \models M_0$. But, if it were so, j should semantically imply some elements of $MVD(U_0)$ which are not deducible from M_0 ; for example, $j \models \emptyset \twoheadrightarrow \{x,y\}$. □

The condition "$J(M) \neq \emptyset$" is characterized in [BFMY] and in [H2]. When $J(M) \neq \emptyset$, we say that M is <u>join-realizable.</u>

By the precedent lemma, we can state the following :

<u>Lemma</u>

For a subset M of MVD(U), the set J(M) has the following properties :

(1) $\forall j \in J(M)\ (j \geq_{JD} M)$,

(2) $\forall j \in J(M)\ (j =_{JD} M \iff \exists M' \subseteq MVD(U)(j =_{JD} M'))$. □

Let us call an element j of JD(U) to be <u>perfect</u>, when there exists a subset M of MVD(U) such that $j =_{JD} M$. The perfectness of j is shown to be a desirable property and characterized by a certain acyclicity of j's hypergraph, in [FMU] and [BFMY]. We shall use the term "acyclic" in the sense of these works. Instead of saying that j's hypergraph is acyclic, we shall say that j is <u>acyclic</u> or the data base schema of j is <u>acyclic</u>.

When J(M) contains perfect elements, we call them <u>perfect join-realizations</u> of M.

Does every join-realizable M have a perfect join-realization ? As counter-example and as an example for other uses, we note the following example.

Example

Let U_1 be a set of five distinct attributes, say $U_1 = \{x,y,z,v,w\}$ and let M_1 be the subset of $MVD(U_1)$ given by : $M_1 = \{xz \twoheadrightarrow y,\ yv \twoheadrightarrow z\}$.

Then M_1 is join-realizable. The elements of $J(M_1)$ are all JD-equivalent ($=_{JD}$) to one of the two JD's, $\bowtie \underline{X}_0$ and $\bowtie \underline{Y}_0$ with $\underline{X}_0$ (see figure 1) and $\underline{Y}_0$ of $C(U_1)$ as follows :

$$\underline{X}_0 = \{xy,yz,zv,vw,wx,xv\},$$

$$\underline{Y}_0 = \{xy,yz,zv,vwx\},$$

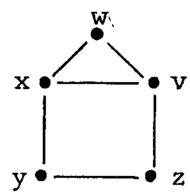

Figure 1

where xy,...,vwx stand for the subsets $\{x,y\},\ldots,\{v,w,x\}$ of U_1, respectively.

We can verify that $\bowtie \underline{X}_0 >_{JD} M_1$ and $\bowtie \underline{Y}_0 >_{JD} M_1$. That is, we have $\forall j \in J(M_1)(j >_{JD} M)$. We may also try to find a subset M' of MVD(U) such that $\bowtie \underline{X}_0 =_{JD} M'$ or that $\bowtie \underline{Y} =_{JD} M'$, to understand finally that it is impossible.

We note also that $\bowtie \underline{X}_0 >_{JD} \bowtie \underline{Y}_0$ and that $\underline{Y}_0$ can be obtained by eliminating a cycle from $\underline{X}_0$ whereas the cycle in $\underline{Y}_0$ can no more be eliminated without influencing the MVD-structure. $\square$

We have seen that not every subset M of MVD(U) has a perfect join-realization. And it seems that the subsets M which have no perfect join-realization, have already in themselves an undesirable structural character. For example, the subset M_1 of $MVD(U_1)$ above has a pair of elements, $xz \twoheadrightarrow y \mid vw$ and $yv \twoheadrightarrow z \mid xw$, crossing each other and cannot be reduced to another equivalent subset having no crossing pair.

We therefore define that a subset M of MVD(U) is <u>simple</u>, when M has a perfect join-realization, i.e. :

$$M \text{ is simple} \iff_{def} \exists j \in J(M)\ (j \text{ is perfect}).$$

By the precedent lemma, if $j \in J(M)$, then the condition "j is perfect" is equivalent to the condition "$j =_{JD} M$". And if $j =_{JD} M$, then $j \in J(M)$ is trivial. Thus we have :

Lemma

For any subset M of MVD(U) :

M is simple $\Leftrightarrow \exists j \in JD(U)(j =_{JD} M)$. □

Now suppose that M is simple. Then J(M) has a perfect element j' by definition, and as we have seen in the above, $j' =_{JD} M$. On the other hand, we have seen that $\forall j \in J(M)(j \geq_{JD} M)$. We have therefore $\forall j \in J(M)(j \geq_{JD} j')$. Recall that j' was taken to be perfect. So j' is acyclic.

The precedent example suggests that if there are two elements j_1 and j_2 in the same set J(M) and if $j_1 \geq_{JD} j_2$ holds, then j_1 can be reduced to j_2 by eliminating some cycles in j_1. That is, the structure of j_1 can be decomposed into two independent parts, i.e. the general structure given by j_2 and the local structures given by the eliminated cycles.

If this is true, in the case of J(M) with a simple M, the structure of each element j of J(M) can be decomposed into the general structure given by the perfect element j', which is therefore acyclic, and the local structures on the components of this acyclic structure, which give no influence on the corresponding MVD-structure. That is, such a j can be treated nearly as an acyclic one in the practice.

So, when an element j of JD(U) has a subset M of MVD(U), which is simple and such that $j =_{MVD} M$, we say also that j is <u>simple</u>. That is, for any element j of JD(U)

j is simple $\Leftrightarrow_{def} \exists M \subseteq MVD(U)(j =_{MVD} M$ and M is simple).

When j is simple, the corresponding data base schema $\hat{j}$ is also called <u>simple</u>.

As we see easily, this notion is characterizable as follows.

Lemma

For any element j of JD(U) :

j is simple <=> $\exists j' \in JD(U)(j =_{MVD} j'$ and j' is perfect). □

Now we want to know efficient ways to determine whether a given element j of JD(U) is simple or not and whether a given join-realizable subset M of MVD(U) is simple or not.

For these purpose, we pay a particular attention to the uppermost elements of concerning J(M).

To connect the given j to such an element of concerning J(M), we introduce an upward inner reduction of every J(M).

To connect the given M to such an element, we have an intermediate object. It is a subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2} = \{B \subseteq U \mid card(B) = card(U)-2\}$, called "basis of M". It is ensured by [H2] that we can associate to every join-realizable M one and only one subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$. To associate an uppermost element of J(M) to the above $\mathcal{B}$, we have a very simple set theoretical function. Note that two subsets M,M' such that $M =_{MVD} M'$ are mapped to the same $\mathcal{B}$, because the relation "$\mathcal{B}$ is a basis of M" is defined in such a way that it is found to be equivalent to a relation of the form :

$$\forall m \in MVD(U)(\mathcal{B} \sqsupset\!\!\sim m \iff M \models m),$$

with a certain relation "$\sqsupset\!\!\sim$" (see [H1]).

It remains to determine whether the uppermost element thus obtained is simple or not. This can be done by a simple criterion whose justification is given by using the result of [FMU] and [BFMY], i.e. characterization of the perfectness by the acyclicity.

Let us begin our study by :

CHARACTERIZING J(M) BY A BASIS OF M

A basis is a sort of filter for MVD-structures, relying on the notion of agreement. For any subset B of U and for any element m of MVD(U), the relation "B agrees with m" denoted by "$B \supset\!\sim m$", is defined as follows :

$$B \supset\!\sim X \twoheadrightarrow Y \iff_{def} X \subseteq B \Rightarrow (Y \subseteq B \vee (U-Y) \subseteq B).$$

This notions is extended to the set level in the following ways :

$$\mathcal{B} \supset\!\sim m \iff_{def} \forall B \in \mathcal{B} (B \supset\!\sim m)$$

$$B \supset\!\sim M \iff_{def} \forall m \in M (B \supset\!\sim m)$$

$$\mathcal{B} \supset\!\sim M \iff_{def} \forall B \in \mathcal{B} \forall m \in M (B \supset\!\sim m).$$

For any subset M of MVD(U), a subset $\mathcal{B}$ of $\mathcal{P}(U)$ is called a basis of M, iff the following condition holds :

$$\forall m \in MVD(U) (\mathcal{B} \supset\!\sim m \iff BASE(M) \supset\!\sim m),$$

where BASE(M) or more precisely BASE<U,MVD>(M) is a subset of $\mathcal{P}(U)$ as follows :

$$BASE\langle U,MVD\rangle(M) =_{def} \{B \in \mathcal{P}(U) \mid B \supset\!\sim M\}.$$

According to [H1], we have the following :

Lemma

$\mathcal{B}$ is a basis of M, iff :

$$\forall m \in MVD(U) (\mathcal{B} \supset\!\sim m \iff M \models m). \qquad \square$$

Let us denote by $\mathcal{P}(U)_{\geq i}$ with $1 \leq i \leq card(U)$, the set of all the subsets of U whose cardinality is $\geq i$, and denote by $\mathcal{P}(U)_i$ the difference set $(\mathcal{P}(U)_{\geq i}) - (\mathcal{P}(U)_{\geq i+1})$. And instead of $\mathcal{P}(U)_{n-i}$ with $n = card(U)$ and with $1 \leq i < n$, let us also use the notation $\mathcal{P}(U)_{-i}$.

According to [H2], all the MVD-structures on U which are join-realizable, can be characterized by the subsets of $\mathcal{P}(U)_{-2}$, that is :

Lemma

For any subset M of MVD(U) :

1) If M is join-realizable, then there is a unique subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ such that $\mathcal{B}$ is a basis of M($\mathcal{B}$ may be empty).

2) If M is not join-realizable, then there is no subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ such that $\mathcal{B}$ is a basis of M.

Proof

See Corollary 2 and Corollary 3 of [H2]. □

By the above two lemmas, we can conclude that every equivalence class w.r.to "$=_{MVD}$" of join-realizable subsets of MVD(U) is characterized by a subset of $\mathcal{P}(U)_{-2}$. Consequently, every equivalence class w.r. to "$=_{MVD}$" of JD's on U, corresponds bijectively to a subset of $\mathcal{P}(U)_{-2}$.

So, let us define for any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ a subset $J(\mathcal{B})$ of JD(U) as follows :

$$J(\mathcal{B}) =_{def} \{j \in JD(U) \mid \forall m \in MVD(U)(j \models m \iff \mathcal{B} \supset\!\sim m)\}.$$

Then we can state this fact as follows.

Lemma

Let j be any element of JD(U). Then there is one and only one subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ such that $[j]_{=MVD} = J(\mathcal{B})$; where $[j]_{=MVD}$ denotes the equivalence class of j :

$$[j]_{=MVD} =_{def} \{j' \in JD(U) \mid j' =_{MVD} j\}.$$

Proof

This is clear by the following sublemma. □

Sublemma

For any $\mathcal{B} \subseteq \mathcal{P}(U)_{-2}$ and for any $M \subseteq MVD(U)$:

$\mathcal{B}$ is a basis of M <=> $J(\mathcal{B}) = J(M)$. □

Now we are interested in :

FINDING A REPRESENTATIVE ELEMENT OF THE CLASS $J(\mathcal{B})$.

Let $\mathcal{B}$ be a subset of $\mathcal{P}(U)_{-2}$. We have formerly noted by an example that some element of $J(\mathcal{B})$ has a form strongly related to the form of $\mathcal{B}$. Let us consider the following element of C(U), that we shall call the induced covering according to $\mathcal{B}$:

$$I(\mathcal{B}) =_{def} \overline{\mathcal{B}} \cup \{\{x\} \mid x \in \cap\mathcal{B}\}$$

where :

$$\overline{\mathcal{B}} =_{def} \{U-B \mid B \in \mathcal{B}\}$$

Clearly the mapping I is bijective from $\mathcal{P}(\mathcal{P}(U)_{-2})$ into $C(U) \cap \mathcal{P}(\mathcal{P}(U)_1 \cup \mathcal{P}(U)_2)$. And for any $\underline{X}$ of the image set $I(\mathcal{P}(\mathcal{P}(U)_{-2}))$, the inverse mapping I^{-1} can be defined as follows :

$$I^{-1}(\underline{X}) =_{def} \overline{(\underline{X} \cap \mathcal{P}(U)_2)}$$

Example

Let U_1 be a set of five distinct attributes, say $U_1 = \{x,y,z,v,w\}$ and let $\mathcal{B}_1 = \{xyw, xzw, xvw, yzw, yvw, zvw\}$. Then :

$$I(\mathcal{B}_1) = \{zv, yv, yz, xv, xz, xy, \{w\}\}$$; see figure 2.

In the above, we have denoted by xyw, zv, etc. the subsets {x,y,w}, {z,v}, etc. of U_1, respectively. We verify also that :

$$I^{-1}(I(\mathcal{B}_1)) = \mathcal{B}_1.$$ □

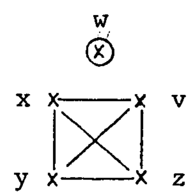

Figure 2

According to [H2], the followings hold.

Lemma

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$\bowtie I(\mathcal{B}) \in J(\mathcal{B})$. □

Corollary

For any element j of JD(U), there exists one and only one subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ such that :

$j =_{MVD} \bowtie I(\mathcal{B})$. □

Example

Let U_1 and $\mathcal{B}_1$ be as the precedent example. Let M_2 be the subset $\{\emptyset \twoheadrightarrow w\}$ of $MVD(U_1)$. Then we can verify that :

$\mathcal{B}_1 \succ m \Longleftrightarrow M_2 \models m$,

and :

$\bowtie I(\mathcal{B}_1) \models m \Longleftrightarrow M_2 \models m$,

for any element m of $MVD(U_1)$. It follows that $\bowtie I(\mathcal{B}_1)$ is an element of $J(\mathcal{B}_1)$. □

Now let us denote by SC(j) with $j \in JD(U)$ and by SC(J) with $J \subseteq JD(U)$, the corresponding element and subset of C(U), respectively. That is, if $j = \bowtie \underline{X}$, then $SC(j) = \underline{X}$ and $SC(J) = \{SC(j) \mid j \in J\}$. For example, $SC(\bowtie I(\mathcal{B})) = I(\mathcal{B})$.

The element $I(\mathcal{B})$ of $SC(J(\mathcal{B}))$ is graph-theoretically defined from $\mathcal{B}$ but $SC(J(\mathcal{B}))$ is not. A graph-theoretical characterization of $SC(J(\mathcal{B}))$ is our nest interest.

For the visual advantage, let us denote by $\hat{j}$ and $\hat{J}(\mathcal{B})$, the element SC(j) and the subset SC(J($\mathcal{B}$)) of C(U), respectively. Our problem is then to find a graph-theoretical relation between $\hat{j}$ and I($\mathcal{B}$), which holds if and only if $j =_{MVD} \bowtie I(\mathcal{B})$.

To help this approach, we insert in the following a graph-theoretical characterization of the order relation "$\models$" on the set JD(U).

A GRAPH-THEORETICAL CHARACTERIZATION OF "$\models$" ON JD(U)

Let $\underline{X}$, $\underline{Y}$ be any pair of elements of C(U). The relation "$\underline{X}$ is finer than $\underline{Y}$", denoted by "$\underline{X} \geq \underline{Y}$" is defined as follows :

$$\underline{X} \geq \underline{Y}$$

$$\Longleftrightarrow_{def} \forall X \in \underline{X}\ \exists Y \in \underline{Y}\ (X \subseteq Y)$$

Clearly this relation satisfies the conditions of order relation, i.e. it is reflexive and transitive. The relation "$\underline{X}$ is equivalent to $\underline{Y}$" and "$\underline{X}$ is strictly finer than $\underline{Y}$", denoted by "$\underline{X} \simeq \underline{Y}$" and by "$\underline{X} > \underline{Y}$", respectively, are defined from this relation as usual.

Theorem

For any elements $\underline{X}$ and $\underline{Y}$ of C(U) :

$$\underline{X} \geq \underline{Y} \Longleftrightarrow \bowtie \underline{X} \models \bowtie \underline{Y}.$$

Proof

=>) Suppose that $\underline{X} \geq \underline{Y}$ and $R \mid\!\sim \bowtie \underline{X}$. We have to show that $R \mid\!\sim \bowtie \underline{Y}$, that is :

$$\forall f \in \text{TUPLE}(U)(\forall Y \in \underline{Y}(f[Y] \in R[Y]) \Rightarrow f \in R).$$

Let f be a tuple on U such that $\forall Y \in \underline{Y}(f[Y] \in R[Y])$. By $R \mid\!\sim \bowtie \underline{X}$, it suffices to deduce from this the condition : $\forall X \in \underline{X}(f[X] \in R[X])$.

Let X be any element of $\underline{X}$. Then by the definition of $\underline{X} \geq \underline{Y}$, there is some element X' of $\underline{Y}$ such that $X \subseteq X'$. X' must satisfy $f[X'] \in R[X']$. Therefore, the subset X of X' satisfy $f[X] \in R[X]$.

<=) We shall show that $\underline{X} \not\leq \underline{Y} \Rightarrow \bowtie \underline{X} \not\models \bowtie \underline{Y}$. Let us fix an element X_0 of $\underline{X}$ such that $\forall Y \in \underline{Y}(X_0 \not\subseteq Y)$, and also fix a pair of tuples f_0, f_1 on U such that $\forall x \in U(f_0(x) \neq f_1(x))$. And we shall construct on this basis a relation R on U such that $R \mid\!\sim \bowtie \underline{X}$ and $R \not\mid\!\sim \bowtie \underline{Y}$.

For every subset W of U, let us denote by f_W the tuple on U such that :

$$\begin{cases} f_W[U-W] = f_0[U-W], \\ f_W[W] = f_1[W]. \end{cases}$$

And let $\underline{Z}$ be a covering of the fixed subset X_0, obtained from X_0 and $\underline{Y}$ as follows :

$$\underline{Z} = \{X_0 \cap Y \mid Y \in \underline{Y} \text{ and } X_0 \cap Y \neq \emptyset\}.$$

Now define a relation R on U by :

$$R = \{f_Z \mid Z \in \underline{Z}\}$$

We first show that $R \mid\!\sim \bowtie \underline{X}$.

Let f be an element of $\bowtie \{R[X] \mid X \in \underline{X}\}$. In order to show that $f \in R$, it suffices showing that :

$$f[X_0] = f_Z[X_0] \Rightarrow \forall X \in \underline{X}(f[X] = f_Z[X]),$$

as far as Z is an element of $\underline{Z}$.

We verify this by the fact that :

$$f[X] = \begin{cases} f_Z[X] & \text{if } X \cap Z \neq \emptyset \\ f_0[X] & \text{otherwise,} \end{cases}$$

and also by the fact that :

$$f_0[X] = f_Z[X], \text{ if } X \cap Z = \emptyset$$

We next show that $R \not\mid\!\sim \bowtie \underline{Y}$.

We want to find a tuple g such that :

$g \in \bowtie \{R[Y] \mid Y \in \underline{Y}\}$ and $g \notin R$.

Take f_{X_0} for g. Then $g \notin R$, because $X_0 \notin \underline{Z}$ by the hypothesis : $\forall Y \in \underline{Y}(X_0 \not\subseteq Y)$. And $g \in \bowtie\{R[Y] \mid \underline{Y} \in Y\}$ is clear, because :

$f_{X_0}[Y] = f_{X_0 \cap Y}[Y]$ for any $Y \in \underline{Y}$. □

This theorem shows that the order relation $j \geq_{JD} j'$ on JD(U) is isomorphically represented by the order relation $\hat{j} \geq \hat{j}'$ on C(U). Consequently, the equivalence relation $j =_{JD} j'$ on JD(U) is also isomorphically represented by $\hat{j} \simeq \hat{j}'$. That is :

<u>Corollary</u>

For any elements j, j' of JD(U)

1) $j \geq_{JD} j' \iff \hat{j} \geq \hat{j}'$,

2) $j =_{JD} j' \iff \hat{j} \simeq \hat{j}'$. □

<u>NORMAL FORMS OF DATA BASE SCHEMAS</u>

By definition of the relation "≥", we have :

$\underline{X} \subseteq \underline{Y} \Rightarrow \underline{X} \geq \underline{Y}$,

in general. But if each element of the difference $\underline{Y} - \underline{X}$ is always upper-bounded by some element of $\underline{X}$, then the inverse relation "$\underline{Y} \geq \underline{X}$" also holds. So it is possible that both "$\underline{X} \subsetneq \underline{Y}$" and "$\underline{X} \simeq \underline{Y}$" hold at the same time. That is, a schema may contain some elements which are redundant. We therefore introduce the <u>normalization reduction</u> whose unit reduction "$:\simeq_1$" is given by the following rule (N).

(N) $\underline{X} :\simeq_1 \underline{X} - \{X\}$,

for any $\underline{X} \in C(U)$ and for any $X \in \underline{X}$ such that $\exists X' \in \underline{X}(X \subsetneq X')$. □

We say that $\underline{X}$ is N-reducible to $\underline{Y}$ and denote it by "$\underline{X} :\simeq \underline{Y}$", if $\underline{Y}$ is obtainable from $\underline{X}$ by successive (may be null) applications of this reduction rule, (N). When a schema $\underline{X}$ does not contain any element X such that $\exists X' \in \underline{X}(X \subsetneq X')$, we say that $\underline{X}$ is normal or of normal form.

The relation "$:\simeq$" satisfies the Church-Rosser property :

$$\exists \underline{X}(\underline{X} :\simeq \underline{Y} \text{ and } \underline{X} :\simeq \underline{Z}) \Rightarrow \exists \underline{W}(\underline{Y} :\simeq \underline{W} \text{ and } \underline{Z} :\simeq \underline{W}).$$

Let us define the relation "$:\simeq:$" by :

$$\underline{X} :\simeq: \underline{Y} \Leftrightarrow_{\text{def}} \exists \underline{Z} \in C(U)(\underline{X} :\simeq \underline{Z} \text{ and } \underline{Y} :\simeq \underline{Z}).$$

Then by the Church-Rosser property of "$:\simeq$", this relation is transitive and therefore is an equivalence relation. As expected, the following holds.

Lemma

For any $\underline{X}$, $\underline{Y}$ of $C(U)$,

$$\underline{X} \simeq \underline{Y} \Leftrightarrow \underline{X} :\simeq: \underline{Y}$$

Proof

$\Rightarrow$) Let $\underline{X} \simeq \underline{Y}$ hold. Then both $\underline{X} \cup \underline{Y} :\simeq \underline{X}$ and $\underline{X} \cup \underline{Y} :\simeq \underline{Y}$ hold. This implies $\underline{X} :\simeq: \underline{Y}$, by the Church-Rosser property of "$:\simeq$".

$\Leftarrow$) Trivial. □

Corollary

Every equivalence class of $C(U)$ with respect to the relation "$\simeq$" has a unique normal element. □

For any element $\underline{X}$ of $C(U)$, the normal element of the equivalence class $[\underline{X}]_{\simeq}$ is called the normal form of $\underline{X}$, denoted by $|X|$. We extend this notation ot a subset S of $C(U)$:

$$|S| =_{def} \{|\underline{X}| \mid \underline{X} \in S\}.$$

It is called the <u>norm set</u> of S.

<u>Example</u>

Let $C(U)_{\leq 2}$ be the subset of $C(U)$ defined by :

$$C(U)_{\leq 2} =_{def} C(U) \cap \mathcal{P}(\mathcal{P}(U)_1 \cup \mathcal{P}(U)_2).$$

Then obviously :

$$|C(U)_{\leq 2}| = \{I(\mathcal{B}) \mid \mathcal{B} \in \mathcal{P}(\mathcal{P}(U)_{=2})\}.$$

Let $\hat{J}(\mathcal{B})_{\leq 2}$ be the intersection of $\hat{J}(\mathcal{B})$ and $C(U)_{\leq 2}$, for every $\mathcal{B} \in \mathcal{P}(\mathcal{P}(U)_{=2})$. Then its norm $|\hat{J}(\mathcal{B})_{\leq 2}|$ is clearly the singleton set $\{I(\mathcal{B})\}$ That is, we can characterize the set $\hat{J}(\mathcal{B})_{\leq 2}$ as follows :

$$\hat{J}(\mathcal{B})_{\leq 2} = \{\underline{X} \in C(U) \mid |\underline{X}| = I(\mathcal{B})\}.$$ □

<u>Note</u>

In this example, we have get a characterization of the uppermost part $\hat{J}(\mathcal{B})_{\leq 2}$ of $\hat{J}(\mathcal{B})$, by means of the normalization reduction. The normalization reduction is based on the equivalence relation "$\simeq$" on $C(U)$, corresponding to the equivalence relation "$=_{JD}$" on $JD(U)$.

To get a characterization of the full set $\hat{J}(\mathcal{B})$, we want to have a reduction on $C(U)$ reducing every element of $\hat{J}(\mathcal{B})$ to an element of $\hat{J}(\mathcal{B})_{\leq 2}$. Such reduction should be based on the equivalence relation on $C(U)$, corresponding to the equivalence relation "$=_{MVD}$" on $JD(U)$.

In the following section, we shall try to characterize the order relation "$\geq_{MVD}$" in view of obtaining the equivalence relation on $C(U)$ mentioned above. □

A GRAPH-THEORETICAL CHARACTERIZATION OF "$\geq_{MVD}$"

For any element $\underline{Z}$ of C(V) with $V \subseteq U$, let us denote by $\pi(\underline{Z})$ the partition of V, consisting of the connected components of $\underline{Z}$. As the set of the partitions $\Pi(V)$ of V is a subset of C(V), π gives a mapping from C(V) onto $\Pi(V) \subseteq C(V)$. We suppose that the relation "$\geq$" is defined on every C(V) in the same way as on C(U), in particular, it is defined on every $\Pi(V)$. The following should be noted.

Lemma

For any elements π_1 and π_2 of $\Pi(V)$:

$$\pi_1 \geq \pi_2 \iff \forall Z \in \pi_2 \exists \pi' \subseteq \pi_1 (Z = \cup\pi'). \qquad \Box$$

That is, $\pi_1 \geq \pi_2$ means that π_1 is finer than π_2. Using these order relations "$\geq$" on every $\Pi(V)$, we define an order relation on C(U), "$\underline{X} \underset{\sim}{\trianglerighteq} \underline{Y}$", reading that $\underline{X}$ is <u>more flexible</u> than $\underline{Y}$, as follows :

$$\underline{X} \underset{\sim}{\trianglerighteq} \underline{Y} \iff_{def} \forall V \subseteq U(\pi(\underline{X}|V) \geq \pi(\underline{Y}|V)),$$

where $\underline{X}|V$ and $\underline{Y}|V$ denote the restriction of $\underline{X}$ and $\underline{Y}$ on V, respectively ; i.e. $\underline{X}|V$, for example, is an element of C(V) defined by :

$$\underline{X}|V =_{def} \{X \cap V \mid X \in \underline{X} \text{ and } X \cap V \neq \emptyset\}.$$

When both $\underline{X} \underset{\sim}{\trianglerighteq} \underline{Y}$ and $\underline{Y} \underset{\sim}{\trianglerighteq} \underline{X}$ hold, we say that $\underline{X}$ <u>has the same flexibility as</u> $\underline{Y}$ and denote it by $\underline{X} \approx \underline{Y}$. When $\underline{X} \underset{\sim}{\trianglerighteq} \underline{Y}$ holds but not $\underline{Y} \underset{\sim}{\trianglerighteq} \underline{X}$, we say that $\underline{X}$ is <u>strictly more flexible</u> than $\underline{Y}$ and denote it by $\underline{X} \underset{\not\sim}{\trianglerighteq} \underline{Y}$. It is clear that the relation "$\underset{\sim}{\trianglerighteq}$" is an order relation on C(U) and the relation "$\approx$" is an equivalence relation on C(U).

Note

$\underline{X} \geq \underline{Y}$ implies $\underline{X} \underset{\sim}{\trianglerighteq} \underline{Y}$ for any $\underline{X}, \underline{Y} \in C(U)$. $\Box$

We intend to show that $\underline{X} \underset{\sim}{\trianglerighteq} \underline{Y}$ is equivalent to $\bowtie \underline{X} \geq_{MVD} \bowtie \underline{Y}$. We shall apply the following theorem due to [FMU] (theorem 3).

<u>Theorem [FMU]</u>

For any $j \in JD(U)$ and for any $V \twoheadrightarrow W \in MVD(U)$, the relation "$j \models V \twoheadrightarrow W$" holds if and only if there exists a subset. π' of $\pi(\hat{j}\,|U-V)$ such that $W-V = \cup\pi'$. □

By this theorem we can establish the following :

<u>Theorem</u>

For any elements $\underline{X}$, $\underline{Y}$ of $C(U)$:

$\underline{X} \trianglerighteq \underline{Y} \iff \bowtie \underline{X} \geq_{MVD} \bowtie \underline{Y}$.

<u>Proof</u>

=>) We suppose $\bowtie \underline{Y} \models V \twoheadrightarrow W$ and we want to prove that $\bowtie \underline{X} \models V \twoheadrightarrow W$, using the condition $\underline{X} \trianglerighteq \underline{Y}$. By theorem [FMU], this reduces to showing the following implication using the condition $\underline{X} \trianglerighteq \underline{Y}$:

$$\exists \pi' \subseteq \pi(\underline{Y}|U-V)(W-V = \cup\pi')$$

$$\Rightarrow \exists \pi'' \subseteq \pi(\underline{X}|U-V)(W-V = \cup\pi'').$$

But by definition of $\underline{X} \trianglerighteq \underline{Y}$, we have $\pi(\underline{X}|U-V) \geq \pi(\underline{Y}|U-V)$, which verifies the above implication, by the precedent lemma.

<=) Let V be any subset of U. We want to show $\pi(\underline{X}|V) \geq \pi(\underline{Y}|V)$, assuming that $\bowtie \underline{X} \geq_{MVD} \underline{Y}$. By assumption, for any subset W of V, the relation $\bowtie \underline{Y} \models U-V \twoheadrightarrow W$ implies the relation $\bowtie \underline{X} \models U-V \twoheadrightarrow W$. By theorem [FMU], this means that :

$$\exists \pi' \subseteq \pi(\underline{Y}|V)(W = \cup\pi')$$

$$\Rightarrow \exists \pi'' \subseteq \pi(\underline{X}|V)\ (W = \cup\pi'').$$

Suppose that $\pi(\underline{Y}|V) = \{W_1, W_2, \ldots, W_n\}$.Then for every W_i $(i=1,2,\ldots,n)$, the relation :

$$\exists \pi' \subseteq \pi(\underline{Y}|V)(W_i = \cup\pi')$$

is trivially satisfied. So, by the above implication, the relation :

$$\exists \pi_i'' \subseteq \pi(\underline{X}|V)(W_i = \cup\pi_i'')$$

holds for every $W_i (i=1,2,\ldots,n)$.

That is, we have obtained that :

$$\forall W \in \pi(\underline{Y}|V) \exists \pi'' \in \pi(\underline{X}|V)(W = \cup\pi'').$$

By the previous lemma, this is equivalent to :

$$\pi(\underline{X}|V) \geq \pi(\underline{Y}|V),$$

what we wanted to show. □

Corollary

For any elements $\underline{X}$, $\underline{X}'$ of C(U),

$$\underline{X} \approx \underline{X}' \iff \bowtie\underline{X} =_{MVD} \bowtie\underline{X}'.$$ □

Corollary

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$$\hat{J}(\mathcal{B}) = \{\underline{X} \in C(U) \mid \underline{X} \approx I(\mathcal{B})\}.$$ □

CHARACTERIZATION OF $\hat{J}(\mathcal{B})$ BY MEANS OF A REDUCTION

We have characterized the relation "$=_{MVD}$" on JD(U) by the relation "$\approx$" on C(U), and the relation "$\geq_{JD}$" on JD(U) by the relation "$\geq$" on C(U). Consulting the definitions of these relations "$\approx$" and "$\geq$", we arrive to formulate a reduction of elements of C(U), which goes upwardsbut along with the equivalence "$\approx$". We shall call it the clique reduction or the cliquetization. Its unit reduction "$:\leq_1$" is given by the following rule (C), where Clique (X) with $X \subseteq U$, denotes the set of subsets of X which have the cardinality 2, i.e. Clique(X) $=_{def} \{Y \in \mathcal{P}(U)_2 \mid Y \subseteq X\}$.

(C) $\underline{X} :\leq_1 (\underline{X} - \{X\}) \cup \mathrm{Clique}(X)$,

for any $\underline{X} \in C(U)$ and for any element X of $\underline{X}$ such that $\mathrm{card}(X) \geq 3$. □

We say that a schema $\underline{X}$ is clique-reducible or C-reducible to a schema $\underline{Y}$, if we can obtain $\underline{Y}$ from $\underline{X}$ by successive (may be null) applications of the rule (C). We denote it by "$\underline{X} :\leq \underline{Y}$". At every unit reduction of the clique-reduction, the number of elements X such that $\mathrm{card}(X) \geq 3$ diminishes by one. So by the finiteness of the number of elements of schemas, the length of the reduction must be finite. A schema for which we can no more apply the unit reduction, is called clique-irreducible. A schema is clique-irreducible if and only if it belongs to $C(U)_{\leq 2}$. Clearly we have :

Lemma

The clique-reducibility ":≤" satisfies the Church-Rosser property. □

So, for any schema $\underline{X}$, there is one and only one schema $\underline{X}'$ such that $\underline{X} :\leq \underline{X}'$ and that $\underline{X}'$ is clique-irreducible. We call it the clique of $\underline{X}$ and denote it by $\mathrm{Clique}(\underline{X})$. As mentioned above, $\mathrm{Clique}(\underline{X}) \in C(U)_{\leq 2}$. And in the previous section, we have seen that $\hat{J}(\mathcal{B})_{\leq 2} = \hat{J}(\mathcal{B}) \cap C(U)_{\leq 2}$ is characterizable as the set of schemas $\underline{X}$ such that $|\underline{X}| = I(\mathcal{B})$. So we ask if the clique-reduction may imply the property that
$\underline{X} \in \hat{J}(\mathcal{B}) \Rightarrow \mathrm{Clique}(\underline{X}) \in \hat{J}(\mathcal{B})_{\leq 2}$.

Lemma

For any elements $\underline{X}$, $\underline{X}'$ of $C(U)$,

$\underline{X} :\leq_1 \underline{X}' \Rightarrow \underline{X} \approx \underline{X}'$

Proof

We have to verify if $\pi(\underline{X}|V) = \pi(\underline{X}'|V)$ holds for any subset V of U. But this is clear, because for any subset X of U and for any pair of elements x,y of V, we have :

x and y are connected by $X \cap V$
<=> x and y are connected by $\mathrm{Clique}(X)|V$. □

Corollary

For any elements $\underline{X}$, $\underline{X}'$ of C(U) :

$$\underline{X} :\leq \underline{X}' \Leftrightarrow \underline{X} \approx \underline{X}'. \qquad \square$$

Corollary

For every subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ and for any element $\underline{X}$ of C(U),

$$\underline{X} \in \hat{J}(\mathcal{B}) \Leftrightarrow \mathrm{Clique}(\underline{X}) \in \hat{J}(\mathcal{B})_{\leq 2}. \qquad \square$$

We have already seen that $\hat{J}(\mathcal{B})_{\leq 2} = \{\underline{X} \in C(U) \mid |\underline{X}| = I(\mathcal{B})\}$. Thus we have get the following characterization of $\hat{J}(\mathcal{B})$.

Theorem

For every subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$$\hat{J}(\mathcal{B}) = \{\underline{X} \in C(U) \mid |\mathrm{Clique}(\underline{X})| = I(\mathcal{B})\}. \qquad \square$$

Corollary

For any elements $\underline{X}$, $\underline{X}'$ of C(U),

$$\underline{X} \approx \underline{X}' \Leftrightarrow |\mathrm{Clique}(\underline{X})| = |\mathrm{Clique}(\underline{X}')|. \qquad \square$$

Example

Let U_1, $\mathcal{B}_1$ and M_2 be as the precedent example. Then all of the following data base schemas are elements of $\hat{J}(\mathcal{B}_1)$:

$\underline{X}_1 = \{xyz, xv, yv, zv, \{w\}\}$,

$\underline{Y}_1 = \{xyz, yzv, xv, \{w\}\}$,

$\underline{Z}_1 = \{xyz, yzv, zvx, \{w\}\}$,

$\underline{V}_1 = \{xyz, yzv, zvx, vxy, \{w\}\}$,

$\underline{W}_1 = \{xyzv, \{w\}\}$.

Note that they are of normal form. Note also that $\bowtie \underline{W}_1 =_{JD} M_2$, i.e. $M_2 \models \bowtie \underline{W}_1$. $\square$

FRAME - ABSTRACTION

The examples we have seen above shows that certain kind of cycles are eliminable from a data base schema without alternating its MVD-structure. By such eliminations of cycles a data base schema is transformed into a schema lowerly situated in the equivalence class, and finally we may get a lowermost schema in the class ; giving the macroscopique view on the structure of the given schema. When the given schema is a simple schema, the lowermost schema thus obtained must be a perfect schema, therefore an acyclic schema.

We want to formulate such a reduction on C(U). In the following we define a reduction, called the frame abstraction or FA-reduction, and show that it satisfies the expected properties.

Let us extend the definition of the relation "$\geq$" on C(U) to $\mathcal{P}(\mathcal{P}(U))$, that is, for any subsets $\mathcal{B}$, $\mathcal{B}'$ of $\mathcal{P}(U)$:

$$\mathcal{B} \geq \mathcal{B}' \iff_{def} \forall X \in \mathcal{B}\ \exists Y \in \mathcal{B}'(X \subseteq Y).$$

The unit reduction "$:>_1$" of FA-reduction is given by the following rule, (FA) :

(FA) $\underline{X} :>_1 (\underline{X} - \mathcal{P}(Y)) \cup \{Y\}$,

for any $\underline{X} \in C(U)$ and for any $Y \subseteq U$ such that :

(1) $\text{card}(Y) \geq 3$,

(2) $\text{Clique}(Y) \geq \underline{X}$,

(3) $\{Y\} \not\geq \underline{X}$.

A schema $\underline{X}$ is said to be FA-reducible to a schema $\underline{Y}$, when $\underline{Y}$ can be obtained from $\underline{X}$ by successive (may be null) applications of the rule (FA).

Pose that $\underline{Z} = (\underline{X} - \mathcal{P}(Y)) \cup \{Y\}$. Then by definition of $\underline{Z}$, it is clear that $\underline{X} \geq \underline{Z}$ holds, and consequently $\underline{X} \underset{\sim}{\trianglerighteq} \underline{Z}$ holds (see the note on the definition of "$\underset{\sim}{\trianglerighteq}$").

On the contrary, the condition (3) prevents that $\underline{Z} \geq \underline{X}$ holds. But the condition (2) functions to maintain the weaker property, $\underline{Z} \underset{\sim}{\trianglerighteq} \underline{X}$, by implying that $\underline{Z}|V \geq \pi(\underline{X}|V)$, for every $V \subseteq U$.

Resuming these, we can state the following.

Lemma

For any elements $\underline{X}$, $\underline{Y}$ of $C(U)$,

$$\underline{X} :>_1 \underline{Y} \quad \Rightarrow \quad \underline{X} \approx \underline{Y} \quad \text{and} \quad \underline{X} > \underline{Y}. \qquad \square$$

By the fact that $\underline{X} :>_1 Y$ implies $\underline{X} > \underline{Y}$, any chain of FA-reduction does not form a cycle. As $C(U)$ contains only finite number of elements, we can state the following.

Corollary

Every chain of FA-reduction terminate by an FA-irreducible schema. $\square$

The condition for the FA-reducibility is considerably weak.

Lemma

For any element $\underline{X}$ of $C(U)$:

$$\exists\, \underline{Z} \in C(U)(\underline{X} \approx \underline{Z} \text{ and } \underline{Z} \not\geq \underline{X}) \Rightarrow \exists\, \underline{Y} \in C(U)(\underline{X} :>_1 \underline{Y}).$$

Proof

Assume that $\underline{X} \approx \underline{Z}$ and $\underline{Z} \not\geq \underline{X}$. These two conditions together show that there is at least one element Y in $\underline{Z}$ such that $\mathrm{card}(Y) \geq 3$, which satisfies $\{Y\} \not\geq \underline{X}$; i.e. Y satisfies the condition (1) and (3) of the rule (FA). Now we want to show that such a Y satisfies also the condition (2) of the rule (FA). In fact, by $Y \in \underline{Z}$, we have $\underline{X} \cup \{Y\} \subseteq \underline{X} \cup \underline{Z}$, which implies $\underline{X} \cup \{Y\} \geq \underline{X} \cup \underline{Z}$, therefore $\underline{X} \cup \{Y\} \underset{\sim}{\trianglerighteq} \underline{X} \cup \underline{Z}$. But by $\underline{X} \approx \underline{Z}$, we have $\underline{X} \cup \underline{Z} \approx \underline{X}$. Combining these two, we have $\underline{X} \cup \{Y\} \underset{\sim}{\trianglerighteq} \underline{X}$, from which we conclude that $\mathrm{Clique}(Y) \geq \underline{X}$. That is, we have proved that there is a subset Y of U which satisfies the three conditions of the rule (FA) with respect to $\underline{X}$. So, $\underline{X} :>_1 \underline{Y}$ with $\underline{Y} = (\underline{X} - \mathcal{P}(Y)) \cup \{Y\}$. $\square$

Noting that the converse of the above lemma follows immediately from the precedent lemma, we can state the following.

Corollary

For any elements $\underline{X}$, $\underline{Y}$ of C(U),

$\underline{X} \approx \underline{Y}$

$\Leftrightarrow \exists \underline{X}', \underline{Y}' \in C(U)(\underline{X} :> \underline{X}' \wedge \underline{Y} :> \underline{Y}' \wedge \underline{X}' \simeq \underline{Y}')$.

Proof

=>) There are chains of FA-reduction terminating by FA-irreducible schemas, from $\underline{X}$ to $\underline{X}'$ and from $\underline{Y}$ to $\underline{Y}'$. By the first lemma, $\underline{X} \approx \underline{X}'$ and $\underline{Y} \approx \underline{Y}'$. But by hypothesis, $\underline{X} \approx \underline{Y}$. Therefore $\underline{X}' \approx \underline{Y}'$. By the second lemma, this implies $\underline{Y}' \geq \underline{X}'$ and $\underline{X}' \geq \underline{Y}'$, recalling that $\underline{X}'$ and $\underline{Y}'$ are FA-irreducible.

<=) By definition, $\underline{X}' \simeq \underline{Y}'$ implies $\underline{X}' \approx \underline{Y}'$. And by the first lemma, $\underline{X} :> \underline{X}'$ and $\underline{Y} :> \underline{Y}'$ imply $\underline{X} \approx \underline{X}'$ and $\underline{Y} \approx Y'$, respectively. These together imply $\underline{X} \approx \underline{Y}$. □

FA-reduction has also the following property.

Lemma

FA-reducibility ":>" satisfy the Church-Rosser property.

Proof

Suppose that $\underline{X} :>_1 \underline{Y}$ and $\underline{X} :>_1 \underline{Z}$. If $\underline{Y} \simeq \underline{Z}$, then $\underline{Y} = \underline{Z}$. If $\underline{Y} > \underline{Z}$, then $\underline{Y} :>_1 \underline{Z}$. If $\underline{Y} \ngeq \underline{Z}$ and $\underline{Z} \ngeq \underline{Y}$, then we can find $\underline{W}$ such that $\underline{Y} :>_1 \underline{W}$ and $\underline{Z} :>_1 \underline{W}$. The verifications are easy. The general case, $\underline{X} :> \underline{Y}$ and $\underline{X} :> \underline{Z}$, is by induction on the sum of the numbers of reduction steps, from $\underline{X}$ to $\underline{Y}$, and from $\underline{X}$ to $\underline{Z}$. □

As a consequence, every schema $\underline{X}$ can be reduced to a unique FA-irreducible schema, which we shall call <u>the frame</u> of $\underline{X}$, denoted by Frame($\underline{X}$).

There are elements in $C(U)-C(U)_{\leq 2}$, which have no predecessor with respect to FA-reduction. Therefore $C(U)$ cannot be generated from $C(U)_{\leq 2}$ by FA-reduction. But the normal part $|C(U)|$ of $C(U)$ can be generated from $|C(U)_{\leq 2}|$ by FA-reduction.

Lemma

For any element $\underline{Y}$ of $|C(U)| - |C(U)_{\leq 2}|$, there exists an element $\underline{X}$ of $|C(U)|$ such that $(\underline{X} :>_1 Y)$.

Proof

Take any element Y of $\underline{Y}$ such that $\mathrm{card}(Y) \geq 3$. Then $\underline{X} = |(\underline{Y} - \{Y\}) \cup \mathrm{Clique}(Y)|$ satisfies $\underline{X} :>_1 \underline{Y}$. The verification is easy. □

Moreover we have :

Lemma

For any element $\underline{X}$ of $|C(U)|$,

$$\forall \underline{Y} \in C(U)(\underline{X} :>_1 \underline{Y} \Rightarrow \underline{Y} \in |C(U)|) \qquad \square$$

These two lemma together imply the following :

Corollary

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$$\hat{J}(\mathcal{B}) \cap |C(U)| = \{\underline{X} \in C(U) \mid I(\mathcal{B}) :> \underline{X}\}. \qquad \square$$

As $\underline{X} \simeq \underline{Y}$ implies $\underline{X} \approx \underline{Y}$, we can generally say that $\underline{X} \approx |\underline{X}|$. So if $\underline{X} \in \hat{J}(\mathcal{B})$, then $|\underline{X}| \in \hat{J}(\mathcal{B})$. From this we can conclude that $|\hat{J}(\mathcal{B})| = \hat{J}(\mathcal{B}) \cap |C(U)|$. So we have :

Theorem

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$$\hat{J}(\mathcal{B}) = \{\underline{X} \in C(U) \mid I(\mathcal{B}) :> |\underline{X}|\}. \qquad \square$$

By the precedent lemma, we know also that Frame(I($\mathcal{B}$)) is normal. Using this fact we get another characterization of $\hat{J}(\mathcal{B})$.

Theorem

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$$\hat{J}(\mathcal{B}) = \{\underline{X} \in C(U) \mid |\mathrm{Frame}(\underline{X})| = \mathrm{Frame}(I(\mathcal{B}))\}. \qquad \square$$

CHARACTERIZATION OF THE SIMPLICITY

The notion of simplicity is defined for three kinds of objects : subsets M of MVD(U), elements j of JD(U) and elements $\underline{X}$ of C(U). In this section, we firstly show that the simplicity of these objects reduces to that of I($\mathcal{B}$) with $\mathcal{B} \in \mathcal{P}(\mathcal{P}(U)_{-2})$ which can be uniquely determined by these objects. We secondly show that the simplicity of $\bowtie$I($\mathcal{B}$) is equivalent to the perfectness of $\bowtie$ Frame(I($\mathcal{B}$)), which implies the equivalence between the simplicity of I($\mathcal{B}$) and the acyclicity of Frame (I($\mathcal{B}$)), by the result of [FMU]. Using this equivalence and using the graph-theoretical relation between I($\mathcal{B}$) and Frame (I($\mathcal{B}$)), we finally give a characterization of the simplicity of I($\mathcal{B}$).

The simplicity of M and the simplicity of I($\mathcal{B}$)

Let M be a subset of MVD(U), supposed to be join-realizable (if not, M cannot be simple by definition). By the results of [H2], M has a basis $\mathcal{B}$ in $\mathcal{P}(\mathcal{P}(U)_{-2})$ and such $\mathcal{B}$ is uniquely determined by M (note that there is even a simple algorithm to compute it from M). And we have shown that J(M) = J($\mathcal{B}$). We have also shown that J($\mathcal{B}$) contains $\bowtie$I($\mathcal{B}$).

Now suppose M is simple. By definition and by the fact that J(M) = J($\mathcal{B}$), this is equivalent to the condition :

$$\exists\, j \in J(\mathcal{B}) \text{ (j is perfect).}$$

But the condition $j \in J(\mathcal{B})$ is equivalent to $j =_{MVD} \bowtie I(\mathcal{B})$.

The above condition is therefore equivalent to say that $\bowtie$I($\mathcal{B}$) is simple. We have establish the following.

Lemma

For any subset M of MVD(U) which is join-realizable, there is one and only one subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$, which is a basis of M and for which the following equivalence holds :

M is simple <=> I($\mathcal{B}$) is simple. □

<u>The simplicity of j and the simplicity of I($\mathcal{B}$)</u> .

Let j be any element of JD(U). We have learned that we can obtain from $\hat{j}$ a uniquely determined element $|\mathrm{Clique}(\hat{j})|$ of C(U), and it should be of the form I($\mathcal{B}$) with some subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$. We know also that $\hat{j}$ and I($\mathcal{B}$) are both in $\hat{J}(\mathcal{B})$. Therefore j and $\bowtie$I($\mathcal{B}$) are both in J($\mathcal{B}$). The simplicity of j and of $\bowtie$I($\mathcal{B}$), means equally the existence of a perfect element in J($\mathcal{B}$). We have established the following.

<u>Lemma</u>

For any element j of JD(U), there is a unique subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ such that $|\mathrm{Clique}(\hat{j})| = I(\mathcal{B})$, for which the following equivalence holds :

j is simple <=> I($\mathcal{B}$) is simple. □

<u>The simplicity of $\underline{X}$ and the simplicity of I($\mathcal{B}$)</u>

In the same way as the case of JD, we have :

<u>Lemma</u>

For any element $\underline{X}$ of C(U), there is a unique subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$ such that $|\mathrm{Clique}(\underline{X})| = I(\mathcal{B})$,for which the following equivalence holds :

$\underline{X}$ is simple <=> I($\mathcal{B}$) is simple. □

<u>The simplicity of $\bowtie$I($\mathcal{B}$) and the perfectness of $\bowtie$Frame(I($\mathcal{B}$))</u>

We show a more general form.

<u>Lemma</u>

For any element $\underline{X}$ of C(U),

$\bowtie\underline{X}$ is simple <=> $\bowtie$Frame($\underline{X}$) is perfect.

Proof

By definition of Frame($\underline{X}$) and by isomorphisms between "$\geq$", "$\simeq$" and "$\geq_{JD}$", "$=_{MVD}$", respectively, it holds :

(1) $\bowtie \underline{X} =_{MVD} \bowtie \mathrm{Frame}(\underline{X})$,

(2) $\forall j \in \mathrm{JD}(U)\ (j =_{MVD} \bowtie \mathrm{Frame}(\underline{X}) \Rightarrow j \geq_{JD} \bowtie \mathrm{Frame}(\underline{X}))$.

The direction ($\Leftarrow$) of the lemma is trivial by (1).

The direction ($\Rightarrow$) of the lemma : Suppose that $\bowtie \underline{X}$ is simple. Then there is an element j' of JD(U) such that $\bowtie \underline{X} =_{MVD} j'$ and that j' is perfect. As j' is perfect, there is a subset M of MVD(U) such that $j' =_{JD} M$. We have to show that $\mathrm{Frame}(\underline{X}) =_{JD} M$. From $\bowtie \underline{X} =_{MVD} j'$ follows $j' \geq_{JD} \bowtie \mathrm{Frame}(\underline{X})$, by (1) and (2), therefore $M \geq_{JD} \bowtie \mathrm{Frame}(\underline{X})$. On the other hand, we see that $\bowtie \mathrm{Frame}(\underline{X})$ is an element of J(M). It follows that $\bowtie \mathrm{Frame}(\underline{X}) \geq_{JD} M$. We have implied that there is an M such that $\bowtie \mathrm{Frame}(\underline{X}) =_{JD} M$. □

The next step is the replacement of the notion of perfectness by that of acyclicity, which we want to precise here in our terminology.

Let $\underline{X}$ be an element of C(U). For every subset V of U, we define the <u>set of the candidates of articulation sets</u> of $\underline{X}|V$ by :

$$\mathrm{CA}(\underline{X}|V) =_{def} \{X \cap Y \mid X, Y \in \underline{X}|V,\ X \cap Y \neq \emptyset,\ X \not\subseteq Y,\ Y \not\subseteq X\}.$$

The set of the articulation sets of $\underline{X}|V$ is the subset of $\mathrm{CA}(\underline{X}|V)$ defined by :

$$A(\underline{X}|V) =_{def} \{Q \in \mathrm{CA}(\underline{X}|V) \mid \#\pi(\underline{X}|V-Q) > \#\pi(\underline{X}|V)\}.$$

We say that $\underline{X}$ is <u>cyclic</u> if $\mathrm{CA}(\underline{X}|V) \neq 0$ and $A(\underline{X}|V) = \emptyset$ for some subset V of U. Otherwise, $\underline{X}$ is said to be <u>acyclic</u>.

Now we state the theorem due to [FMU].

Theorem [FMU]

For any element $\underline{X}$ of $C(U)$,

$\underline{X}$ is acyclic $\Leftrightarrow \bowtie \underline{X}$ is perfect. □

Combining this with the precedent lemma, we have :

Lemma

For any element $\underline{X}$ of $C(U)$,

$\underline{X}$ is simple $\Leftrightarrow$ Frame($\underline{X}$) is acyclic. □

Another lemma which may be useful is the following :

Lemma

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$$\text{Clique}(\text{Frame}(I(\mathcal{B}))) = I(\mathcal{B}).$$

Proof

As Frame($I(\mathcal{B})$) is an element of $\hat{J}(\mathcal{B})$,

$$|\text{Clique}(\text{Frame}(I(\mathcal{B})))| = I(\mathcal{B}).$$

On the other hand, Frame($I(\mathcal{B})$) is normal. Therefore :

$$|\text{Clique}(\text{Frame}(I(\mathcal{B})))| = \text{Clique}(\text{Frame}(I(\mathcal{B}))).$$ □

The next step is the characterization of the simplicity of $I(\mathcal{B})$. We may consider $I(\mathcal{B})$ as a graph on U in the usual sense.

For any subset G of $\mathcal{P}(U)$, the union $\cup G$ is called the <u>support</u> of G and denoted by SUPP(G) or by $\bar{G}$. We recall that $I(\mathcal{B})$ with $\mathcal{B} \in \mathcal{P}(\mathcal{P}(U)_{-2})$, is a normal element of $C(U)_{\leq 2}$. We denote by $I(\mathcal{B})_1$ and $I(\mathcal{B})_2$ the subsets of $I(\mathcal{B})$, $I(\mathcal{B}) \cap \mathcal{P}(U)_1$ and $I(\mathcal{B}) \cap \mathcal{P}(U)_2$, respectively. By the normality, SUPP($I(\mathcal{B})_1$) and SUPP($I(\mathcal{B})_2$) are disjoint. $I(\mathcal{B})$ can be represented by the graph $(U, I(\mathcal{B})_2)$.

A subset C of $I(\mathcal{B})_2$ is called a <u>cycle</u> of $I(\mathcal{B})$, if for any element x of $SUPP(C)$, there are exactly two elements of C that contain x. A cycle C of $I(\mathcal{B})$ is said to be a <u>pure cycle with respect to</u> $I(\mathcal{B})$, if $Clique(\bar{C}) - C \neq \emptyset$ and $(Clique(\bar{C}) - C) \cap I(\mathcal{B})_2 = \emptyset$. We say that $I(\mathcal{B})$ <u>has a pure cycle</u>, if there is a cycle of $I(\mathcal{B})$ which is a pure cycle with respect to $I(\mathcal{B})$.

What we want to show is the following.

<u>Lemma</u>

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$I(\mathcal{B})$ has a pure cycle <=> $Frame(I(\mathcal{B}))$ is cyclic.

<u>Proof</u> : Let us denote by $F(\mathcal{B})$ the set $Frame(I(\mathcal{B}))$.

=>) Suppose that C is a pure cycle w.r.t. $I(\mathcal{B})$. Put $V = SUPP(C)$. Then $F(\mathcal{B})|V = C$. Hence $CA(F(\mathcal{B})|V) = SUPP(C) \neq \emptyset$ and $A(F(\mathcal{B})|V) = \emptyset$. That is, $F(\mathcal{B})$ is cyclic.

<=) Suppose that $F(\mathcal{B})$ is cyclic. That is, there is a subset V of U such that $CA(F(\mathcal{B})|V) \neq \emptyset$ and $A(F(\mathcal{B})|V) = \emptyset$.

Take a maximal element Q of $CA(F(\mathcal{B})|V)$ w.r.t. the inclusion. By definition of $CA(F(\mathcal{B})|V)$, there must be found in $F(\mathcal{B})|V$ at least a pair of supersets X, Y of Q, such that $X \not\subseteq Y$ and $Y \not\subseteq X$. Without loss of generality, we may suppose that they are maximal in $F(\mathcal{B})|V$ w.r.t. the inclusion. By the maximality of Q, we have $X \cap Y = Q$ (So $X-Q$ and $Y-Q$ are disjoint).

Q cannot be an element of $A(F(\mathcal{B})|V)$, by hypothesis. It means that the disjoint non-empty sets $X-Q$ and $Y-Q$ should be contained in a common component of the partition $\pi(F(\mathcal{B})|V-Q)$ of $V-Q$, because X and Y are contained in a common component of the partition $\pi(F(\mathcal{B})|V)$. That is, $X-Q$ and $Y-Q$ should be connected in $F(\mathcal{B})|V-Q$, and therefore also in $I(\mathcal{B})_2|V-Q$.

Take a path of the minimum length in $I(\mathcal{B})_2|V-Q$, from $X-Q$ to $Y-Q$. Let $z_o, z_1, z_2, \ldots, z_n$ be the corresponding sequence of points, with $z_o = x \in X-Q$, $z_n = y \in Y-Q$ and $z_1, z_2, \ldots, z_{n-1} \in V-(X \cup Y)$.

There must exist in Q an element q for which $z_1 q$ does not belong to $I(\mathcal{B})|V$. Because, if it were not so, $\text{Clique}(Q \cup \{x,z_1\}) \subseteq I(\mathcal{B})|V$, implying that $F(\mathcal{B})|V$ has an element Z such that $Q \cup \{x,z_1\} \subseteq Z$, which is contradictory in both of the following cases.

Case 1. $X \subsetneq Z$. This contradicts to the maximality of X.

Case 2. $X \not\subseteq Z$. Then the intersection $Q' = X \cap Z$ is strictly greater than Q, i.e. $Q \subsetneq Q'$, which contradicts to the maximality of Q.

The existence of such q, implies also that $z_1 \neq y$.

We next examine whether $z_2 q$ is in $I(\mathcal{B})|V$. If not, $z_2 \neq y$ and z_3 exists. In this way,we continue to examine successively until we arrive to find z_k such that $z_k q$ is in $I(\mathcal{B})|V$ and that $z_i q$ with $1 \leq i < k$ are all not in $I(\mathcal{B})|V$.

The cycle C corresponding to the sequence $z_o, z_1, z_2, \ldots, z_k, q, z_o$ is a pure cycle w.r.t. $I(\mathcal{B})|V$, because the path was taken to be of the minimum length.

It remains to verify that C is a pure cycle w.r.t. $I(\mathcal{B})$. For this it suffices to show that

$$(\text{Clique}(\bar{C})-C) \cap I(\mathcal{B})_2 \subseteq (\text{Clique}(\bar{C})-C) \cap I(\mathcal{B})_2\, V.$$

But this can be verified by the relation

$$\text{Clique}(\bar{C}) \subseteq \text{Clique}(V),$$

which can be reduced to the fact that C is a subset of $I(\mathcal{B})_2|V$.

Now we can state the following.

Lemma

For any subset $\mathcal{B}$ of $\mathcal{P}(U)_{-2}$,

$I(\mathcal{B})$ is simple <=> $I(\mathcal{B})$ has no pure cycle. □

We have established the following.

Theorem

1) For any subset M of MVD(U), M is simple if and only if there is a basis $\mathcal{B}$ of M in $\mathcal{P}(\mathcal{P}(U)_{-2})$ such that $I(\mathcal{B})$ has no pure cycle.

2) For any element j of JD(U), j is simple if and only if $|\mathrm{Clique}(\hat{j})|$ has no pure cycle. □

CONCLUSION

The notion of simplicity is introduced as an extension of the notion of acyclicity. It is introduced not only for a single JD j but also for a set M of MVDs.

The simplicity applied to j characterizes the property that j can be represented by an acyclic JD j' accompanied by a set of embedded JDs on the components of j', such that the suppression of each of these embedded JDs does not influence the MVD-structure implied by j, which is perfectly representable by the acyclic JD j', that is, the simplicity of j means that j can be decomposed into two mutually orthogonal parts : the global constraint represented by an acyclic JD j' and the local constraint represented by a set of embedded JDs on the components of j'. When j is not simple, we cannot obtain the orthogonality. In the case of a set M of MVDs, the simplicity means only the perfect representability of M by an acyclic JD. In this case, the second part of the orthogonal decomposition does not appear.

Because of the orthogonality in the decomposition and because of the acyclicity of the first part of the decomposition, the notion of simplicity may find applications in the data base design and in the method of the legality check.

Our result can be also applied as an algorithm of determining the acyclicity.

As a tool of our study we have used the notion of bases for MVD-structure. Applying this notion to the mixed FD + MVD-structures, we may obtain a similar result on the mixed FD + JD-structures.

REFERENCES

[A] W.W. ARMSTRONG
"Dependency structures of data base relationships"
in INFORMATION PROCESSING 74, North-Holland Publ. CO. 1974.

[AD] W.W. ARMSTRONG, C. DELOBEL
"Decompositions and functional dependencies in relations"
ACM Transactions on Database Systems, Vol. 5, N°4, Dec. 1980.

[BFMY] C. BERRI, R. FAGIN, D. MAIER, M. YANNAKAKIS
"On the desirability of acyclic database schemes"
Research Report IBM, RJ 3131, 1981.

[FMU] R. FAGIN, A.O. MENDELZON, J.D. ULLMAN
"A simplified universal relation assumption and its properties"
Research Report IBM, RJ 2900, 1980.

[H1] Y. HANATANI
"Bases for FD-MVD-structures", to appear (Rapport de Recherche INRIA).

[H2] Y. HANATANI
"Graph-theoretical classification and generation of JD's by means of bases for MVD-structures", to appear (Rapport de Recherche INRIA).

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique